

**Directivas e categorias para
identificação e classificação semântica
na colecção dourada do HAREM**

Nuno Cardoso e Diana Santos

DI-FCUL

TR-06-18

Departamento de Informática
Faculdade de Ciências da Universidade de Lisboa
Campo Grande, 1749-016 Lisboa
Portugal

Technical reports are available at <http://www.di.fc.ul.pt/tech-reports>. The files are stored in PDF, with the report number as filename. Alternatively, reports are available by post from the above address.

Directivas e categorias para identificação e classificação semântica na colecção dourada do HAREM

Nuno Cardoso[†] e Diana Santos[‡]

[†]Departamento de Informática, Faculdade de Ciências da Universidade de Lisboa

[‡]SINTEF ICT, Oslo

[†]ncardoso@xldb.di.fc.ul.pt, [‡]diana.santos@sintef.no

Resumo

Neste relatório técnico apresentam-se as directivas usadas na compilação da colecção dourada do HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas (REM) em português, organizada pela Linguatca. A colecção dourada (dois conjuntos, de 129 e 128 textos de vários géneros literários) foi manualmente anotada com a identificação de nomes próprios e a sua classificação semântica, com base num leque de categorias derivado quer do consenso inicial entre os participantes no HAREM quer do estudo aturado subsequente das realidades nos textos que analisávamos. Cedo se tornou claro que a documentação das categorias empregues, a sua delimitação o mais precisa possível e a criação de critérios claros para anotação de mais texto em português seriam uma das contribuições mais importantes deste esforço de anotação.

O presente documento inclui pois uma relação extensa dos vários casos encontrados em texto em português e de quais os critérios usados na sua anotação, critérios esses que foram usados na criação da colecção dourada, com base na qual os sistemas participantes foram avaliados no HAREM. O resultado desse processo de reflexão e explicitação das opções tomadas tem no entanto interesse para além da própria avaliação de sistemas de reconhecimento de entidades mencionadas, por ser um início de descrição semântica do português baseada em corpora. Visto que o próprio recurso, a colecção dourada, se encontra disponível, o conhecimento das directivas precisas usadas para o criar permite a futuros investigadores levarem a cabo estudos quantitativos e exploratórios do texto usado, assim como alterações sistemáticas das suas opções se com elas discordarem.

O relatório começa por relatar como foi feita a delimitação das entidades a classificar, seguindo-se uma descrição detalhada de quais os critérios para distinguir entre os vários tipos das dez categorias utilizadas no HAREM (PESSOA, ORGANIZAÇÃO, TEMPO, ACONTECIMENTO, COISA, LOCAL, OBRA, ABSTRACÇÃO, VALOR e VARIADO), com copiosos exemplos e descrição de casos problemáticos ou simplesmente potencialmente complexos.

Nota Preliminar

O HAREM constituiu a primeira avaliação (conjunta) para sistemas de reconhecimento de entidades mencionadas (REM) em textos em português [1, 5, 4, 10, 2, 8, 13, 6, 12, 3], no âmbito da responsabilidade da Linguatca de organizar avaliações conjuntas para a comunidade científica interessada no processamento computacional do português [7]. O HAREM foca a tarefa de identificação e classificação de entidades mencionadas (EM, ou seja, nomes próprios) no texto, uma tarefa relevante para várias áreas de processamento de linguagem natural (PLN) como a resposta automática a perguntas, a extracção de informação e a tradução automática, entre outras.

No HAREM procurámos desenvolver uma nova metodologia de avaliação em REM que abrangesse as especificidades da tarefa, tendo contemplado questões que não tinham sido ainda abordadas com

profundidade suficiente pelos anteriores eventos de avaliação internacionais. Exemplos são a vagueza das EM, a caracterização morfológica das mesmas, e a avaliação de expressões apenas parcialmente identificadas.

O HAREM culminou na organização de dois eventos de avaliação, o principal em Fevereiro de 2005 e a sua sequência, o MiniHAREM, em Abril de 2006 [11], e contou com 10 sistemas participantes oriundos de 6 países (Brasil, Dinamarca, Espanha, França, México e Portugal). Os participantes enviaram um total de 38 saídas, ou seja, anotações automáticas da colecção de textos utilizando os seus sistemas REM, que foram avaliadas. Os resultados foram publicados, e os relatórios detalhados do desempenho de cada sistema foram entregues aos respectivos participantes.

A 15 de Julho de 2006, a Linguatca organizou o Encontro do HAREM na Universidade do Porto, logo a seguir à Primeira Escola de Verão da Linguatca, que reuniu os participantes, organizadores e outros interessados no HAREM. No encontro, os participantes apresentaram os seus sistemas, a organização apresentou detalhadamente o seu trabalho, e todos debateram várias questões sobre o futuro do HAREM, numa sessão final, que demonstrou bem o interesse da comunidade em prosseguir com mais eventos de avaliação em REM. Está assim em curso a organização de um livro (electrónico) sobre o HAREM [9] englobando as comunicações nesse encontro e outra documentação relevante.

Os participantes no HAREM tiveram um papel activo no desenvolvimento da metodologia e na anotação das colecções de texto segundo a metodologia aprovada. Adicionalmente, a organização do HAREM desenvolveu uma plataforma de avaliação de sistemas de REM, para aferir o desempenho dos sistemas participantes, que se encontra publicamente disponível no sítio do HAREM, <http://www.linguatca.pt/HAREM>. Outra das contribuições do HAREM que reputamos valiosa foi a criação manual de uma colecção dourada para avaliação, juntamente com documentação extensa das directivas usadas, constituindo esses textos a primeira obra dedicada ao tratamento exaustivo em corpora da semântica dos nomes próprios em português.

É essa documentação que pretendemos tornar mais acessível agora, na forma de um par de relatórios técnicos, este e o seu gémeo [4], cuja forma final foi publicada na Web em finais de Março de 2006 por ocasião do MiniHAREM, cristalizando definitivamente as normas usadas no primeiro HAREM.

O trabalho de organização do HAREM enquadra-se no projecto da Linguatca, financiada pela Fundação para a Ciência e Tecnologia (FCT) através do projecto POSI/PLP/43931/2001, e co-financiada pelo POSI.

1 Introdução

Neste documento, apresentamos as directivas usadas na etiquetagem da colecção dourada da primeira edição do HAREM e do mini-HAREM, e, consequentemente, qual o comportamento esperado pelos sistemas que nele participem.

Começamos por descrever o formato do que consideramos um texto anotado com entidades mencionadas (EM), e qual a definição operacional destas. Depois, para cada categoria, explicamos o significado atribuído e detalhamos a sua subcategorização.

Noutro texto [4] será indicada a metodologia seguida para a anotação morfológica da colecção dourada.

2 Regras gerais de etiquetagem

Cada EM é rotulada por uma etiqueta de abertura e uma etiqueta de fecho, cujo formato é semelhante ao das etiquetas usadas em XML. A etiqueta de abertura contém a categoria atribuída, e possui atributos como o tipo ou a classificação morfológica. Na etiqueta de fecho, coloca-se a categoria usada na etiqueta de abertura. Um exemplo de uma EM etiquetada é:

os <PESSOA TIPO="GRUPOMEMBRO" MORF="M,S">Beatles</PESSOA>.

Os nomes das categorias e dos tipos não devem incluir caracteres com acentos e/ou cedilhas, e devem estar em maiúsculas. Ou seja, deverá ser usado <ORGANIZACAO> em vez de <ORGANIZAÇÃO>.

Os valores dos atributos TIPO e MORF devem ser rodeados por aspas.

Não deve haver nenhum espaço imediatamente a seguir à etiqueta de abertura e antes da etiqueta de fecho.

Certo: O <PESSOA TIPO="INDIVIDUAL">João</PESSOA> é um professor.

Errado: O<PESSOA TIPO="INDIVIDUAL"> João</PESSOA> é um professor.

Errado: O <PESSOA TIPO="INDIVIDUAL">João </PESSOA>é um professor.

Se a EM contém espaços, esses devem manter-se inalterados.

Certo: O <PESSOA TIPO="INDIVIDUAL">João Mendes</PESSOA> é um professor.

Errado: O <PESSOA TIPO="INDIVIDUAL">JoãoMendes</PESSOA> é um professor.

As aspas, parênteses, pelicas ou travessões não são para incluir na etiqueta, se englobarem a EM como um todo (ver caso 1). No entanto, são para incluir, caso apenas se apliquem a partes da EM (caso 2) ou façam parte integrante da mesma.

Caso 1

Certo: A ``<OBRA TIPO="ARTE">Mona Lisa</OBRA>''

Errado: A <OBRA TIPO="ARTE">``Mona Lisa''</OBRA>

Caso 2

Certo: O <PESSOA TIPO="INDIVIDUAL">Mike ``Iron'' Tyson</PESSOA>

Certo: <PESSOA TIPO="INDIVIDUAL">John (Jack) Reagan</PESSOA>

Certo: Os resultados foram semelhantes aos produzidos por Diana Santos e colegas <OBRA TIPO="PUBLICACAO">(Santos et al, 2005)</OBRA>.

2.1 Recursividade das etiquetas

Não se deve usar etiquetas dentro de etiquetas, como nos exemplos (errados) seguintes:

Errado: <PESSOA TIPO="GRUPO"><ORGANIZACAO TIPO="SUB">Bombeiros
</ORGANIZACAO></PESSOA>

Errado: <ORGANIZACAO TIPO="INSTITUICAO">Departamento de <ABSTRACCAO
TIPO="DISCIPLINA">Informática</ABSTRACCAO> do IST</ORGANIZACAO>

2.2 Vagueza na classificação semântica

No caso de haver dúvidas entre várias categorias ou tipos, deve utilizar-se o operador “|”. Por exemplo, em “Ajudem os Bombeiros!”, se se considerar que não existe razão para preferir uma das duas seguintes classificações para *Bombeiros*, nomeadamente entre <PESSOA TIPO="GRUPO"> e <ORGANIZACAO TIPO="INSTITUICAO">, devem-se colocar ambas:

Certo: Ajudem os <PESSOA|ORGANIZACAO TIPO="GRUPO|INSTITUICAO">
Bombeiros</PESSOA|ORGANIZACAO>!

Podem ser especificados mais do que uma categoria ou tipo, ou seja, <A|B|C|. . .>.

Caso a dúvida seja entre tipos, deve-se repetir a categoria. Por exemplo, se se estiver em dúvida sobre qual o tipo de organização (EMPRESA ou INSTITUICAO?) na frase “O ISR trata dessa papelada”, deve-se repetir a categoria ORGANIZACAO tantas vezes quantos os tipos indicados:

Certo: O <ORGANIZACAO|ORGANIZACAO TIPO="EMPRESA|INSTITUICAO">ISR
</ORGANIZACAO|ORGANIZACAO> trata dessa papelada.

2.3 Vagueza na identificação

Se houver dúvidas (ou análises alternativas) de qual a identificação da(s) EM(s) que deverá ser considerada correcta, as várias alternativas são marcadas entre as etiquetas <ALT> e </ALT>, que delimitam e juntam as várias alternativas, que são separadas pelo carácter '|'. O exemplo abaixo mostra a etiquetagem a usar, quando não se consegue decidir por uma única identificação:

O <ALT><PESSOA TIPO="GRUPOMEMBRO">Governo de Cavaco Silva</PESSOA>
| Governo de <PESSOA TIPO="INDIVIDUAL">Cavaco Silva</PESSOA></ALT>

2.4 Critérios de identificação de uma EM

Uma EM deve conter pelo menos uma letra em maiúsculas, e/ou algarismos.

Certo: <TEMPO TIPO="DATA">Agosto</TEMPO>

Errado: <TEMPO TIPO="DATA">ontem de manhã</TEMPO>

A única excepção a esta regra abrange os nomes dos meses, que devem ser considerados EM, ou parte de EM, mesmo se grafados com minúscula. Esta excepção deve-se ao facto de haver grafia maiúscula em Portugal e minúscula no Brasil nesse caso.

Certo: <TEMPO TIPO="DATA">agosto de 2001</TEMPO>

Existe também um conjunto de palavras relativas a certos domínios que também são excepções a esta regra, e que serão descritas em detalhe abaixo.

Se uma determinada EM, etiquetada como tal, aparecer depois sem maiúsculas no mesmo texto ou noutra, não deve ser outra vez etiquetada, ou seja, uma EM tem de conter obrigatoriamente pelo menos uma letra maiúscula e/ou algarismos.

No entanto, o inverso não é verdade, isto é, uma palavra com pelo menos uma letra maiúscula ou um número pode não ser uma EM. Um caso clássico são as palavras que iniciam as frases, mas também há que considerar o uso excessivo de maiúsculas em certos géneros de textos, como a *web*, onde casos como *Contactos*, *História*, *Página Inicial*, *Voltar*, *Menu*, *E-mail*, entre outros, não devem ser por regra identificados como EM.

Aplicando o mesmo raciocínio, as frases totalmente escritas em maiúsculas (como acontece em títulos de destaque) deverão ser analisadas cuidadosamente, e só deverão conter etiquetas as EM claras. Por exemplo, se uma linha reza “CLIQUE AQUI PARA VER A EDUCAÇÃO EM 1993”, EDUCAÇÃO não deve ser considerada uma EM, uma vez que, naquele contexto, a palavra não deveria conter nenhuma maiúscula. No entanto, o ano deve ser marcado como TEMPO, de tipo DATA ou PERIODO.

Outro exemplo: “ABALO EM LISBOA SEM VÍTIMAS”. Neste caso, consideramos correcto marcar LISBOA como EM, visto que assumimos que manteria a maiúscula se a frase não fosse exclusivamente grafada em maiúsculas. Note-se, de qualquer maneira, que estes casos caem um pouco fora do âmbito do HAREM, em que se utilizou um critério predominantemente gráfico, baseado nas convenções da língua escrita.

Palavras que foram incorrectamente grafadas apenas com minúsculas não são classificadas pelo HAREM como EM em caso nenhum.

2.5 Relação entre a classificação e a identificação

Embora a classificação deva ter em conta o significado da EM no texto, a identificação (ou seja a sua delimitação) deve restringir-se às regras das maiúsculas enunciadas acima. Ou seja, apenas a parte associada ao nome próprio deve ser identificada, embora classificada, se for caso disso, a entidade maior em que se enquadra. Vejam-se os seguintes exemplos:

Certo: a filha de <PESSOA TIPO="INDIVIDUAL">Giuteyte</PESSOA>
Certo: o tratado de <ACONTECIMENTO TIPO="EFEMERIDE">Tordesilhas
</ACONTECIMENTO> dividiu o mundo

Embora apenas *Tordesilhas* tenha sido identificado, é o tratado de Tordesilhas que é classificado como um ACONTECIMENTO.

Isso também se aplica aos casos em que no texto um fragmento ou parte da EM é compreendida como relatando anaforicamente a uma entidade não expressa na sua totalidade. Por exemplo, na frase “A Revolução de 1930 foi sangrenta, e a de 1932 ainda mais”, deve marcar-se 1932 como <ACONTECIMENTO TIPO="EFEMERIDE"> e não como <TEMPO TIPO="DATA">.

Nos casos em que há enganos de ortografia ou grafia no texto, em particular quando uma palavra tem uma maiúscula a mais ou a menos e tal é notório, escolhemos corrigir mentalmente a grafia (maiúscula /minúscula) de forma a poder classificar correctamente. Além disso, estamos a pensar em marcar estes casos, na colecção dourada, com uma classificação META="ERRO".

Certo: O grupo terrorista <PESSOA TIPO="GRUPO" META="ERRO">Setembro
negro</PESSOA>

Outras excepções, mais sistematicamente apresentadas, são as seguintes:

Para poder distinguir mais facilmente os casos de classes de objectos cujo nome inclui um nome próprio (geralmente de uma pessoa), adicionámos a seguinte regra de identificação para a categoria COISA: a preposição anterior também deve fazer parte da EM em *constante de Planck*, *bola de Berlim* ou *porcelana de Limoges*.

Por outro lado, consideramos que as EM de categoria VALOR e do tipo QUANTIDADE ou MOEDA devem incluir a unidade, independentemente de esta ser grafada em maiúscula ou minúscula.

Finalmente, no caso de doenças, formas de tratamento e certo tipo de acontecimentos consideramos aceitáveis um conjunto finito de nomes comuns precedendo a própria EM, cuja lista exhaustiva se encontrará num apêndice futuro.

2.6 Escolha da EM máxima

Para evitar uma excessiva proliferação de EM com identificações alternativas, os sistemas e CD são construídos de forma a escolher a EM máxima, ou seja, aquela que contém, numa única interpretação possível, o maior número de palavras. Assim, e muito embora fosse possível ter tomado a decisão inversa e pedir, por exemplo, o máximo número de EM com uma interpretação possível separada, a escolha recaiu em preferir a EM maior.

Por exemplo:

Certo: O <PESSOA TIPO="CARGO">ministro dos Negócios Estrangeiros do
Governo Sócrates</PESSOA>
Certo: <ORGANIZACAO TIPO="INSTITUICAO">Comissão de Trabalhadores da
IBM Portugal</ORGANIZACAO>
Certo: <ACONTECIMENTO TIPO="EFEMERIDE">Jogos Olímpicos de Inverno de
2006</ACONTECIMENTO>

As únicas exceções a esta regra são períodos descritos por duas datas, e intervalos de valores descritos por duas quantidades.

3 Categoria PESSOA

3.1 Tipo INDIVIDUAL

Títulos que precedem nomes

Os títulos (*dr.*, *eng.*, *arq.*, *Pe.*, *etc.*) usados no tratamento de uma pessoa devem ser incluídos na EM que delimita essa pessoa.

Formas de tratamento normalmente usadas para anteceder um nome, como presidente, ministro, etc. também devem ser incluídos, assim como graus de parentesco (*tia*, *irmão*, *avó*, etc) quando fazem parte da forma de tratamento. Outras relações profissionais como patrão, chefe, etc. não devem ser incluídos, nem profissões que não façam parte da forma de tratamento.

Certo: O <PESSOA TIPO="INDIVIDUAL">Dr. Sampaio</PESSOA>.
Certo: O <PESSOA TIPO="INDIVIDUAL">presidente Jorge Sampaio</PESSOA>.
Certo: O <PESSOA TIPO="INDIVIDUAL">padre Melícias</PESSOA>.
Certo: O <PESSOA TIPO="INDIVIDUAL">tio Zeca</PESSOA>.
Certo: O acordeonista <PESSOA TIPO="INDIVIDUAL">Miguel Sá</PESSOA>.
Errado: O <PESSOA TIPO="INDIVIDUAL">acordeonista Miguel Sá</PESSOA>.

Cargos incluídos

Os cargos que não estejam separados por uma vírgula do nome devem ser incluídos no tipo INDIVIDUAL. Se houver vírgula, ficam de fora.

Certo: O <PESSOA TIPO="INDIVIDUAL">Presidente da República Jorge Sampaio</PESSOA>, disse...
Certo: O <PESSOA TIPO="CARGO">Presidente da República</PESSOA>, <PESSOA TIPO="INDIVIDUAL">Jorge Sampaio</PESSOA>, disse...

Caso o cargo seja descrito após o nome, aplica-se a mesma regra.

Certo: <PESSOA TIPO="INDIVIDUAL">Jorge Sampaio</PESSOA>, <PESSOA TIPO="CARGO">Presidente da República</PESSOA>, assinou...

Outros

Diminutivos, alcunhas, iniciais, nomes mitológicos e entidades religiosas são etiquetados nesta categoria.

Certo: <PESSOA TIPO="INDIVIDUAL">Zé</PESSOA>.
Certo: <PESSOA TIPO="INDIVIDUAL">' ' Iron ' ' Tyson</PESSOA>.
Certo: <PESSOA TIPO="INDIVIDUAL">John (Jack) Reagan</PESSOA>.
Certo: <PESSOA TIPO="INDIVIDUAL">JFK</PESSOA>.
Certo: <PESSOA TIPO="INDIVIDUAL">Deus</PESSOA>.

EM que não são cargos, mas que referem uma pessoa individual, são para ser etiquetados como tal.

Certo: <PESSOA TIPO="INDIVIDUAL">Vossa Excia</PESSOA>

3.2 Tipo GRUPOIND

Esta categoria representa grupo de indivíduos (do tipo INDIVIDUAL) que não têm um nome “estático” como grupo (ao contrário dos Beatles, por exemplo).

Certo: <PESSOA TIPO="GRUPOIND">Vossas Excias</PESSOA>.
Certo: O <PESSOA TIPO="GRUPOIND">Governo Clinton</PESSOA> foi a...
Certo: Foi em casa dos <PESSOA TIPO="GRUPOIND">Mirandas</PESSOA>.
Certo: O governo de <PESSOA TIPO="GRUPOIND">Cavaco Silva</PESSOA> esteve presente na cerimónia.

No caso de haver um grupo de pessoas discriminadas, deve-se etiquetar cada um dos nomes em separado. Na frase de exemplo “Os tenistas Carlos Guerra e António Gomes foram a Wimbledon”:

Certo: Os tenistas <PESSOA TIPO="INDIVIDUAL">Carlos Guerra</PESSOA> e <PESSOA TIPO="INDIVIDUAL">António Gomes</PESSOA> foram a Wimbledon.
Errado: Os tenistas <PESSOA TIPO="GRUPO">Carlos Guerra e António Gomes</PESSOA> foram a Wimbledon.

3.3 Tipo CARGO

O tipo CARGO deve ser usado na referência de um posto que é ocupado por uma pessoa, mas que poderá no futuro ser ocupado por outros indivíduos. Ou seja, num dado contexto, CARGO pode representar uma pessoa em concreto, mas através da referência ao seu cargo. Note-se que noutros casos a mesma EM (que anotamos de qualquer maneira sempre da mesma forma, como <PESSOA TIPO="CARGO">) pode referir-se ao próprio cargo, que pode ser desempenhado por diferentes pessoas ao longo do tempo. Exemplos: *Papa, Ministro dos Negócios Estrangeiros, Rainha da Abissínia*.

Cargo associado a uma organização

Cargos que possuem na descrição uma organização, devem ter apenas uma etiqueta <PESSOA TIPO="CARGO"> que abrange a organização.

Certo: O <PESSOA TIPO="CARGO">Presidente da ONU</PESSOA> foi...
Errado: O <PESSOA TIPO="CARGO">Presidente</PESSOA> da <ORGANIZACAO>ONU</ORGANIZACAO> foi...

3.4 Tipo GRUPOCARGO

O tipo GRUPOCARGO é análogo ao GRUPOIND, designando EM que referem um conjunto de pessoas, através de um cargo.

Certo: os <PESSOA TIPO="GRUPOCARGO">Ministros dos Negócios Estrangeiros da União Europeia</PESSOA>

3.5 Tipo GRUPOMEMBRO

Este tipo, que apenas não se chama GRUPO para salientar a sua relação com o tipo MEMBRO, abrange EM que se referem a um conjunto de pessoas como membros de uma organização ou conceito semelhante (tal como equipa ou seita).

Certo: Os <PESSOA TIPO="GRUPOMEMBRO">Mórmons</PESSOA> acreditam no profeta John Smith.

Certo: Os <PESSOA TIPO="GRUPOMEMBRO">Genesis</PESSOA> deram um concerto ontem.

Certo: O <PESSOA TIPO="GRUPOMEMBRO">BE</PESSOA> reuniu-se ontem.

Certo: O <PESSOA TIPO="GRUPOMEMBRO">FC Porto</PESSOA> jogou muito bem e venceu o jogo.

Errado: O <ORGANIZACAO>FC Porto</ORGANIZACAO> jogou muito bem e venceu o jogo.

Certo: O <ORGANIZACAO>FC Porto</ORGANIZACAO> tem um estádio...

Errado: O <PESSOA TIPO="GRUPOMEMBRO">FC Porto</PESSOA> tem um estádio.

3.6 Tipo MEMBRO

Nos casos onde um indivíduo é mencionado pela organização que representa (e não um grupo), é marcado com o tipo MEMBRO.

Certo: Ele foi abordado por um <PESSOA TIPO="MEMBRO">GNR</PESSOA> à paisana.

Certo: O <PESSOA TIPO="MEMBRO">Mórmon</PESSOA> estava na sala ao lado.

No caso de entrevistas, quando o entrevistador é referenciado pelo nome da publicação, deve ser etiquetado como <PESSOA TIPO="MEMBRO">:

Certo: <PESSOA TIPO="MEMBRO">Jornal Nacional</PESSOA> - O que sente depois de ganhar o prémio?

Errado: <ORGANIZACAO TIPO="EMPRESA">Jornal Nacional</ORGANIZACAO> - O que sente depois de ganhar o prémio?

Os próprios nomes são ABSTRACCAO

Quando o texto foca o nome e não a referência do próprio nome, esse nome (independentemente de se referir a uma pessoa, animal, organização, etc.) é marcado como <ABSTRACCAO TIPO="NOME">:

Certo: Dei-lhe o nome de <ABSTRACCAO TIPO="NOME">João Sem Medo</ABSTRACCAO>.

Errado: Dei-lhe o nome de <PESSOA TIPO="INDIVIDUAL">João Sem Medo</PESSOA>.

Certo: Uma organização suspeita denominada <ABSTRACCAO TIPO="NOME">Os Inimigos das Formigas</ABSTRACCAO> foi ilegalizada ontem no Cairo.

4 Categoria ORGANIZACAO

4.1 Tipo ADMINISTRACAO

Este tipo pretende etiquetar as organizações relacionadas com a administração e governação de um território, tal como ministérios, municípios, câmaras, autarquias, secretarias de estado (Exemplos: *Secretaria de Estado da Cultura, Brasil, Prefeitura de São Paulo, Câmara Municipal de Leiria*). Inclui também as organizações que têm a ver com a governação a nível internacional ou supra-nacional (Exemplos: *ONU, UE*, etc.)

Países ou territórios como organização

EM referentes a países, territórios, regiões autónomas ou mesmo territórios ocupados ou ex-colónias, podem referir à organização, pelo que se deve usar as etiquetas <LOCAL TIPO="ADMINISTRATIVO"> ou <ORGANIZACAO TIPO="ADMINISTRACAO">, dependendo do contexto.

Certo: <ORGANIZACAO TIPO="ADMINISTRACAO">Moçambique</ORGANIZACAO>
votou a favor na ONU.

Certo: <LOCAL TIPO="ADMINISTRATIVO">Moçambique</ORGANIZACAO> faz
fronteira com a Tanzânia.

Referências a ministérios

A referência à entidade organizativa deve ser explícita, para ser considerada uma EM de categoria ORGANIZACAO. No caso em que se refere a uma área de competência da organização, é uma ABSTRACCAO de tipo DISCIPLINA e não uma ORGANIZACAO.

Certo: O <ORGANIZACAO TIPO="ADMINISTRACAO">Ministério do Ambiente
</ORGANIZACAO> gere a política ambiental.

Certo: O ministro do <ABSTRACCAO TIPO="DISCIPLINA">Ambiente
</ABSTRACCAO> gere a política ambiental.

No seguinte caso, a vagueza da EM deve ser mantida (ORGANIZAÇÃO ou ABSTRACCAO)

Certo: O <ORGANIZACAO|ABSTRACCAO TIPO="ADMINISTRACAO|DISCIPLINA">
Ambiente</ORGANIZACAO|ABSTRACCAO> gere a política ambiental.

No caso da menção a Ministro com maiúscula, ambas as situações serão consideradas correctas.

Certo: O <PESSOA TIPO="CARGO">Ministro do Ambiente</PESSOA>
gere a política ambiental.

Certo: O ministro do <ABSTRACCAO TIPO="DISCIPLINA">Ambiente</ABSTRACCAO>
gere a política ambiental.

Esta excepção tem a haver com o facto de escrever Ministro com maiúscula, no contexto apresentado, está errado. As novas versões da colecção dourada estão etiquetadas de maneira a suportar erros como este, sem penalizar os sistemas.

4.2 Tipo EMPRESA

Este tipo abrange organizações com fins lucrativos, como empresas, sociedades, clubes, etc. (Exemplos: *Boavista FC*, *Círculo de Leitores*, *Livraria Barata*, (discoteca) *Kapital*) em contextos em que são mencionadas como tal.

Certo: O <ORGANIZACAO TIPO="EMPRESA">Boavista FC</ORGANIZACAO>
contratou novos jogadores.

4.3 Tipo INSTITUICAO

Todas as organizações que não possuem fins lucrativos (não sendo, portanto, empresas), nem um papel directo na governação, são do tipo INSTITUICAO. Este tipo abrange instituições no sentido estrito, associações e outras organizações de espírito cooperativo, universidades, colectividades, escolas ou partidos políticos (Exemplos: *Associação de Amizade Portugal-Bulgária*, *Universidade Federal do Rio Grande do Sul*, *Liceu Maria Amália*, *PC do B (Partido Comunista do Brasil)*, *Museu do Ar*, *PSP*, *Amnistia Internacional*).

4.4 Tipo SUB

As EM de tipo SUB referem-se a determinados sectores de uma organização, mas sem autonomia para ser considerada ela própria uma organização, tais como departamentos, secções, assembleias gerais, comissões, comités, secretarias, etc.

Certo: A sua queixa deve dirigir-se ao <ORGANIZACAO TIPO="SUB">Departamento dos Alunos de Mestrado do IST</ORGANIZACAO>

Certo: A <ORGANIZACAO TIPO="SUB">Assembleia Geral da Empresa PTO</ORGANIZACAO> tem poder para reprovare o orçamento proposto.

Nome da empresa incluído no SUB

No caso do nome da organização ser parte integrante do tipo SUB, este também faz parte da EM.

Certo: O <ORGANIZACAO TIPO="SUB">Departamento de Marketing da General Motors</ORGANIZACAO>.

Errado: O <ORGANIZACAO TIPO="SUB">Departamento de Marketing da <ORGANIZACAO TIPO=EMPRESA">General Motors </ORGANIZACAO></ORGANIZACAO>.

Errado: O <ORGANIZACAO TIPO="SUB">Departamento de Marketing </ORGANIZACAO> da <ORGANIZACAO TIPO=EMPRESA">General Motors</ORGANIZACAO>.

Sucursais e filiais

No caso de sucursais, filiais, empresas em regime de *franchising*, etc, ou seja, onde haja autonomia suficiente para as considerarmos uma organização autónoma, a EM deve ser classificada como uma EMPRESA, e não uma SUB.

Certo: A <ORGANIZACAO TIPO=EMPRESA">VW Portugal</ORGANIZACAO> vai lançar uma iniciativa.

Errado: A <ORGANIZACAO TIPO=SUB">VW Portugal</ORGANIZACAO> vai lançar uma iniciativa.

Certo: A <ORGANIZACAO TIPO=EMPRESA">GM</ORGANIZACAO> disse à <ORGANIZACAO TIPO=EMPRESA">GM Portugal</ORGANIZACAO> para recolher veículos.

Embora a organização designada pela segunda EM tenha uma dependência explícita em relação à designada pela primeira EM, a sua identificação sai do âmbito do tipo SUB, que pretende delimitar apenas EM que são sectores dentro de uma organização.

Como tal, resumindo:

Certo: <ORGANIZACAO TIPO="EMPRESA">GM Portugal</ORGANIZACAO>

Certo: <ORGANIZACAO TIPO="SUB">Departamento de Vendas da GM</ORGANIZACAO>

Certo: A <ORGANIZACAO TIPO="INSTITUICAO">Faculdade de Ciências da Universidade de Lisboa</ORGANIZACAO>

Certo: A <ORGANIZACAO TIPO="INSTITUICAO">Universidade de Lisboa </ORGANIZACAO> recomendou à <ORGANIZACAO TIPO="INSTITUICAO">

Faculdade de Ciências</ORGANIZACAO>
Certo: A <ORGANIZACAO TIPO="SUB">Reprografia da Universidade de
Lisboa</ORGANIZACAO> fecha às 16h.

Organizações dentro de cargos

Empresas incluídas na descrição dos cargos de pessoas não são para etiquetar.

Certo: O <PESSOA TIPO="CARGO">CEO da Microsoft</PESSOA> foi a...
Errado: o <PESSOA TIPO="CARGO">CEO</PESSOA> da <ORGANIZACAO>
Microsoft</ORGANIZACAO> foi a...

5 Categoria TEMPO

As EM de tipo TEMPO não devem conter palavras que não referem explicitamente a data ou a hora.

Textos como *final de 1999*, *próximo dia 22*, *entre 14 e 18*, *meados de Agosto*, ou *antes do dia 3*, só devem ter marcadas como EM, respectivamente, *1999*, *22*, *14*, *18*, *Agosto* e *3*.

A única excepção é para nomes de meses em português do Brasil, como já foi referido.

Note-se que, embora a idade de uma pessoa seja referida em anos (e, como tal, uma quantidade de tempo), deve ser marcada como <VALOR TIPO="QUANTIDADE"> e não como <TEMPO>, uma vez que se refere a uma quantidade e não a uma localização temporal.

5.1 Tipo DATA

Referência a uma data

Inclui todas as referências a dias, mês e ano. Referências a mês e ano, ou só a ano, devem ser consideradas de tipo DATA se, no contexto, a referência indica uma localização temporal única. Esta pode ter diferentes granularidades (pode ser um dia ou vários meses).

Certo: Camões morreu em <TEMPO TIPO="DATA">1580</TEMPO>.
Certo: O EURO foi em <TEMPO TIPO="DATA">2004</TEMPO>.
Certo: No dia <TEMPO TIPO="DATA">24 de Agosto de 1976</TEMPO>.
Certo: Em <TEMPO TIPO="DATA">Agosto de 1976</TEMPO> foi a Final da Taça.
Errado: Em <TEMPO TIPO="DATA">Agosto de 1976</TEMPO> houve 54 suicídios.
Certo: Em <TEMPO TIPO="PERIODO">Agosto de 1976</TEMPO> houve 54 suicídios.
Certo: Em <TEMPO TIPO="DATA">1974</TEMPO> houve a Revolução.
Errado: Em <TEMPO TIPO="DATA">1974</TEMPO> vendeu-se 200.000 carros.
Certo: Em <TEMPO TIPO="PERIODO">1974</TEMPO> vendeu-se 200.000 carros.

Referência a duas datas

Referências a períodos de tempo através da data de início e da data do final, devem ser etiquetadas com duas EM <TEMPO TIPO="DATA"> separadas, e não com uma única etiqueta <TEMPO TIPO="PERIODO">.

Certo: Entre <TEMPO TIPO="DATA">4</TEMPO> a <TEMPO TIPO="DATA">6 de
Dezembro</TEMPO> há o Festival.
Errado: Entre <TEMPO TIPO="PERIODO">4 a 6 de Dezembro</TEMPO> há o
Festival.

Neste caso, e apenas neste caso, aplicamos a regra de duas EM mínimas são melhores do que uma EM mais longa.

5.2 Tipo HORA

Referência a horas

Aplicam-se as mesmas regras descritas no tipo DATA, para as horas.

Certo: Às <TEMPO TIPO="HORA">2h00</TEMPO> vou ao dentista.
Certo: Entre as <TEMPO TIPO="HORA">2h00</TEMPO> e as <TEMPO TIPO="HORA">4h00</TEMPO> estou no dentista.

Referência a fusos horários

Horas com modificação referente a fusos horários devem abranger essa informação, uma vez que é parte essencial para interpretar o tempo da ocorrência.

Certo: O atentado ocorreu às <TEMPO TIPO="HORA">13h, hora de Lisboa</TEMPO>, e fez...

5.3 Tipo PERIODO

Engloba as EM que referem um intervalo de tempo contínuo e não repetido, com apenas um início e um fim (Exemplos: *Inverno, anos 80, século XIX, 1984, pós-25 de Abril, a Idade do Bronze*). Note-se que a mesma EM pode referir um período único ou cíclico, ou uma data:

Certo: Vou três vezes a Londres no próximo <TEMPO TIPO="PERIODO">Inverno</TEMPO>.
Certo: O <TEMPO TIPO="CICLICO">Inverno</TEMPO> em Oslo costuma ser frio.
Certo: A Joana nasceu no <TEMPO TIPO="DATA">Inverno</TEMPO> passado.

Período referido como um acontecimento

É normal referir um determinado período de tempo através de um evento que decorreu durante esse período. Um exemplo é a *Segunda Guerra Mundial*, que pode ser referenciada como o evento ou como um período de tempo, sendo imprescindível a análise do contexto da EM para definir a semântica correcta, como se mostra nos seguintes exemplos:

Certo : Durante a <TEMPO TIPO="PERIODO">2ª Guerra Mundial</TEMPO>, surgiram os primeiros aviões a jacto.
Certo : A <ACONTECIMENTO TIPO="EFEMERIDE">2ª Guerra Mundial</ACONTECIMENTO> envolveu meio mundo.

A diferença é marcada pela expressão *Durante*, que desde logo indica que a EM é para ser interpretada como um PERIODO.

Certo: Durante a <TEMPO TIPO="PERIODO">Guerra Fria</TEMPO> não houve ataques nucleares.

Período implícito

Semelhante ao caso descrito acima, há outras referências a períodos que são implícitos a partir de diversas EM que, à primeira vista, parecem pertencer a outras categorias semânticas, como é ilustrado abaixo. Por exemplo, tome-se o caso de alguém que trabalhou na IBM e depois passou a trabalhar para a Sun. A sua menção à IBM na seguinte frase refere-se ao período no qual esteve lá empregado.

Certo: Depois da <TEMPO TIPO="PERIODO">IBM</TEMPO>, fui trabalhar para a <ORGANIZACAO TIPO="EMPRESA">Sun</ORGANIZACAO>

Da mesma forma, no contexto de um pessoa que foi trabalhador no navio D. Luís, este deve ser anotado como TEMPO.

Certo: Depois do <TEMPO TIPO="PERIODO">D. Luís</TEMPO>, fiquei desempregado.

5.4 Tipo CICLICO

Compreende períodos recorrentes, quando empregues como tal (*Natal, 1º de Janeiro, Páscoa*).

Há que ter atenção que uma dada EM da categoria TEMPO pode ter quase sempre duas interpretações:

No dia 6 de Novembro comemora-se...

No dia 6 de Novembro vai haver uma greve...

No primeiro caso, como acontece todos os anos, é <TEMPO TIPO="CICLICO">. No segundo caso, é <TEMPO<TEMPO TIPO="DATA">, porque se refere a um único dia.

6 Categoria ACONTECIMENTO

Esta categoria abrange acontecimentos que são únicos e, de uma maneira geral, irrepitíveis (EFEMERIDE), e outros cujo significado se reduz a designar um certo conjunto de actividades e de acções: ORGANIZADO (com sub-partes) e EVENTO (indivisível).

6.1 Tipo EFEMERIDE

Acontecimento ocorrido no passado e não repetível tal como o *25 de Abril*, o *11 de Setembro*, a *2ª Guerra Mundial*.

Certo: A <ACONTECIMENTO TIPO="EFEMERIDE">Revolução Francesa </ACONTECIMENTO> mudou a Europa.

Certo: O <ACONTECIMENTO TIPO="EFEMERIDE"> caso Whitaker </ACONTECIMENTO> abalou a Grã-Bretanha.

6.2 Tipo ORGANIZADO

Acontecimento multifacetado, que poderá durar vários dias, e geralmente conter vários EVENTO. Exemplos são a *Copa*, o *Euro 2004*, os *Jogos Olímpicos*, o *Festival de Jazz do Estoril*.

Acontecimentos periódicos

Quando o acontecimento em questão é um evento periódico, distinguido pelo ano do acontecimento ou pelo seu local, estes (data ou local) devem ser incluídos na etiqueta de acontecimento.

Certo: <ACONTECIMENTO TIPO="ORGANIZADO">Jogos Olímpicos de 2004</ACONTECIMENTO>

Certo: <ACONTECIMENTO TIPO="ORGANIZADO">Jogos Olímpicos de Atenas</ACONTECIMENTO>

Errado: <ACONTECIMENTO TIPO="ORGANIZADO">Jogos Olímpicos</ACONTECIMENTO> de <TEMPO TIPO="DATA">2004</TEMPO>

Errado: <ACONTECIMENTO TIPO="ORGANIZADO">Jogos Olímpicos</ACONTECIMENTO> de <LOCAL TIPO="ADMINISTRATIVO">Atenas</LOCAL>.

6.3 Tipo EVENTO

Acontecimento pontual, organizado ou não, tal como *Benfica-Sporting*, *Chico Buarque no Coliseu*, *Buzinão na Ponte*.

Diferenças entre ORGANIZADO e EVENTO

Um bom exemplo da separação entre os tipos ORGANIZADO e EVENTO é o Euro'2004, que foi um acontecimento ORGANIZADO, que incluiu vários EVENTOS (jogos, festas, conferências, etc).

Quando se diz que um evento pode ser organizado ou não, dá-se o exemplo de um jogo de futebol (organizado) ou de uma manifestação popular espontânea (não organizada).

Note-se o caso apresentado, *Chico Buarque no Coliseu*, onde a combinação de uma PESSOA num determinado LOCAL produz um EVENTO, e como tal, deve ser etiquetado como tal, e não como duas EM distintas.

7 Categoria COISA

Esta categoria abrange coisas podem ser únicas e referenciadas como um item (OBJECTO), podem referir substâncias sem forma ou feitio determinado (SUBSTANCIA), podem representar uma categoria específica que descreve uma população de objectos (CLASSE), ou pode abranger EM cujo significado é um conjunto de objectos, discriminados a partir de uma propriedade comum, e que instancia uma classe de objectos (MEMBROCLASSE)

7.1 Tipo OBJECTO

Refere um objecto ou construção em particular, referido por um nome próprio. Inclui planetas, estrelas, cometas e sóis. Também pode conter objectos específicos.

Certo: A fragata <COISA TIPO="OBJECTO">D. Luís</COISA> atracou ontem.

Certo: Comprámos uma casa ao pé do chalé <COISA TIPO="OBJECTO">Sonho Perfeito</COISA>.

Certo: Consegue-se ver <COISA TIPO="OBJECTO">Marte</COISA> hoje à noite.

7.2 Tipo SUBSTANCIA

Refere substâncias elementares que não se podem considerar objectos, por não serem contáveis (por exemplo, *Paracetamol*, H_2O).

Certo: O <COISA TIPO="SUBSTANCIA">DNA</COISA> é um poço de enigmas.

Certo: O médico disse que tenho falta de vitamina <COISA
TIPO="SUBSTANCIA">D</COISA>.

7.3 Tipo CLASSE

Este tipo, que, convém salientar, junto com MEMBROCLASSE, é análogo à distinção feita na categoria PESSOA entre GRUPOMEMBRO e MEMBRO (CLASSE \iff GRUPOMEMBRO e MEMBROCLASSE \iff MEMBRO), representa classes de objectos que têm um nome e, como tal, dão origem a uma EM (Exemplos: *contador Geiger*, *flauta de Biesel*, *PC*, *SCSI*, *PDF*).

Certo: A FCCN exige relatórios em folhas <COISA TIPO="CLASSE">A4</COISA>.

Certo: Os móveis <COISA TIPO="CLASSE">Luís XV</COISA> são muito raros.

Muitas vezes, uma EM deste tipo refere o 'inventor' da classe (exemplo: *lâmpada de Edison*). O determinante deve ser incluído, para enfatizar essa semântica.

Certo: pêndulo <COISA TIPO="CLASSE">de Foucault</COISA>.

Errado: pêndulo de <COISA TIPO="CLASSE">Foucault</COISA>.

“Consumíveis” tais como *pastéis de Belém*, *bolas de Berlim*, *Tiramisu de chocolate*, *vinho de Setúbal* (num contexto de tipo de vinho, e não como oriundo de um local) também são para ser etiquetados como <COISA TIPO="CLASSE"> (ou tipo SUBSTANCIA, dependendo de serem contáveis ou não).

Certo: Receitas de <COISA TIPO="CLASSE">Bacalhau à Brás</COISA>.

Certo: Adoro bolas <COISA TIPO="CLASSE">de Berlim</COISA>.

Certo: Os pastéis <COISA TIPO="CLASSE">de Belém</COISA> têm muita fama.

7.4 Tipo MEMBROCLASSE

Este tipo abrange EM referentes a uma instanciação de classes, ou seja, objectos determinados que são referidos através da classe a que pertencem. Inclui produtos comercializados, e que são referidos por uma marca ou por uma empresa.

No exemplo “Eu gosto de comer Corn Flakes ao pequeno almoço”, estamos a referir-nos a uma série de produtos comerciais que representam uma classe de objectos (neste caso, cereais de pequeno almoço), mas quando dizemos “Os Corn Flakes de hoje estavam horríveis” referimo-nos a uma instância particular.

Certo: O meu <COISA TIPO="MEMBROCLASSE">Fiat Punto<COISA> foi
à revisão.

Certo: O <COISA TIPO="MEMBROCLASSE">MS Word 2003</COISA> da
Cristina rebentou hoje.

Note-se que, nos casos seguintes, estamos a referir-nos à CLASSE e não a um membro.

Certo : As consolas <COISA TIPO="CLASSE" MORF="F,P">Mega
Drive</COISA> são compatíveis com ...

Certo : Os <COISA TIPO="CLASSE" MORF="M,P">Fiat Punto</COISA>
têm bons travões.

8 Categoria LOCAL

A categoria LOCAL abrange todas as referências a sítios específicos.

8.1 Tipo CORREIO

O tipo CORREIO abrange todas as referências a locais com indicações completas, tais como moradas, números de salas, salas de cinema (Exemplos: *Sala 6*, *Caixa Postal 2400*, *Rua da Escola 15B*). Referências que não incluam endereços completos, ou cuja intenção não é facultar uma morada completa, devem ser do tipo ALARGADO e não CORREIO.

Abrangência do tipo CORREIO

Ao assinalar um <LOCAL TIPO="CORREIO">, deve-se incluir todos os locais inerentes à referência da localização exacta.

Certo: <LOCAL TIPO="CORREIO">Rua Augusta, nº 5 - Lisboa</LOCAL>.

Errado: <LOCAL TIPO="CORREIO">Rua Augusta, nº 5</LOCAL> - <LOCAL TIPO="ADMINISTRATIVO">Lisboa</LOCAL>

8.2 Tipo ADMINISTRATIVO

Identifica localizações que foram criadas e/ou delimitadas pelo Homem. Inclui países, bairros, regiões geopolíticas, entre outros. Exemplos: *Rio de Janeiro*, *Alentejo*, *Bairro dos Anjos*, *Ásia Menor*, *Região Autónoma dos Açores*, *Jardim das Amoreiras*, *Médio Oriente*, *América Latina*, *África*, *Países de Leste*.

Não incluir os prefixos

Não se deve incluir a referência ao tipo de local, caso haja, como são os exemplos distrito, concelho, aldeia, vila, cidade, bairro, região, etc (excepção feita se estas referências tiverem pelo menos uma letra maiúscula).

Certo: vou para a cidade de <LOCAL TIPO="ADMINISTRATIVO">Viseu</LOCAL>.

Errado: vou para a <LOCAL TIPO="ADMINISTRATIVO">cidade de Viseu</LOCAL>.

Certo: vou para a <LOCAL TIPO="ADMINISTRATIVO">Cidade de Viseu</LOCAL>.

Locais dentro de organizações

Não há necessidade de colocar um <LOCAL TIPO="ADMINISTRATIVO"> dentro de ORGANIZACAO. Aliás, já tornámos explícito, acima, que não deve haver encaixe de EM dentro de EM.

Certo: <ORGANIZACAO>Câmara Municipal de Braga</ORGANIZACAO>.

Errado: <ORGANIZACAO>Câmara Municipal de <LOCAL>Braga</LOCAL>
</ORGANIZACAO>.

'Locais' referidos como administração

Nomes de países, cidades, entre outros, designam locais... mas há casos em que a referência ao local é implícita ao seu Governo, ou seja, uma EM de categoria ORGANIZACAO e de tipo ADMINISTRACAO.

Certo: <ORGANIZACAO TIPO="ADMINISTRACAO">Portugal</ORGANIZACAO>
condenou a acção da <ORGANIZACAO TIPO="ADMINISTRACAO">
Indonésia</ORGANIZACAO>.

Errado: <LOCAL TIPO="ADMINISTRATIVO">Portugal</LOCAL> condenou a
acção da <LOCAL TIPO="ADMINISTRATIVO">Indonésia</LOCAL>.

8.3 Tipo GEOGRAFICO

Indica localizações de geografia física que apenas foram baptizadas (e não construídas) pelo Homem.

Não incluir os prefixos

Não se deve incluir o tipo de acidente geográfico, ou seja, referências como rio, serra, mar, península, entre outras, exceptuando se estas contiverem pelo menos uma letra maiúscula.

Certo: Vou ao estuário do <LOCAL TIPO="GEOGRAFICO">Douro</LOCAL>.

Certo: Vou ao estuário do rio <LOCAL TIPO="GEOGRAFICO">Douro</LOCAL>.

Errado: Velejo no <LOCAL TIPO="GEOGRAFICO">rio Douro</LOCAL>.

Certo: Velejo no <LOCAL TIPO="GEOGRAFICO">Rio Douro</LOCAL>.

8.4 Tipo VIRTUAL

O tipo VIRTUAL engloba locais como a Internet, e números de telefone ou de fax, etc., desde que contenham ou algarismos ou letras maiúsculas. URLs ou endereços de correio electrónico não são nunca considerados como EM. Também abrange locais de publicação, referidos pelos nomes dos meios de comunicação social.

Não incluir os prefixos

Só se deve etiquetar os números de telefone, nunca o que os precede!

Certo: Vê o meu sítio na <LOCAL TIPO="VIRTUAL">Internet</LOCAL>.

Certo: Tel: <LOCAL TIPO="VIRTUAL">(096) 347845 4563</LOCAL>.

Errado: <LOCAL TIPO="VIRTUAL">Tel: (096) 347845 4563</LOCAL>.

Referência a local de publicação

Quando o local referido é um sítio abstracto que pode não corresponder a um local físico (como é exemplo um programa ou uma série de televisão ou de rádio), ou é mencionado na função de 'alojamento' de um item (como uma notícia de um jornal), a EM deve ser classificada como do tipo VIRTUAL:

Certo: Podes ler o meu artigo no <LOCAL TIPO="VIRTUAL">Jornal de
Notícias</LOCAL>.

Certo: No <LOCAL TIPO="VIRTUAL">Diário de Notícias</LOCAL> de
hoje, vinha referido...

Neste último caso, a interpretação certa é o local onde estão as notícias (que neste caso, pode ou não corresponder a um suporte de papel). No entanto, é o local que se pretende referenciar, e não o OBJECTO, MARCA ou EMPRESA.

8.5 Tipo ALARGADO

Deve conter referências a locais que não estão nas categorias acima, mas que referem um determinado sítio físico, como é o exemplo de pontos de encontro em edifícios, bares, hotéis, praças, centros de congressos, restaurantes, etc. (Exemplo: *Centro Comercial Amoreiras*).

'Organizações' referidas como LOCAL

Frequentemente, hotéis e centros de congressos são referenciados como sítios de ponto de encontro ou onde ocorrem eventos. Neste caso, nesse contexto essas EM são LOCAL de tipo ALARGADO.

Certo: O Congresso decorrerá no <LOCAL TIPO="ALARGADO">Hotel Beta
</LOCAL> e durará...
Errado: O Congresso decorrerá no <ORGANIZACAO TIPO="EMPRESA">Hotel Beta
</ORGANIZACAO> e durará...
Certo: O <ORGANIZACAO TIPO="EMPRESA">Hotel Beta</ORGANIZACAO> emprega
500 funcionários...
Errado: O <LOCAL TIPO="ALARGADO">Hotel Beta</LOCAL> emprega 500
funcionários...

Esta regra, aliás, aplica-se a todas as EM originalmente de outras categorias, sempre que no contexto remetam para um local concreto, como no seguinte exemplo de um ponto de encontro:

Certo : Encontramo-nos debaixo da <LOCAL TIPO="ALARGADO">Torre Eiffel
</LOCAL>.

Diferença entre ALARGADO e CORREIO

No caso de se referir uma rua, avenida ou praça como um local onde ocorreu ou está localizada qualquer coisa, mas não como se de uma morada ou endereço se tratasse, é um <LOCAL TIPO="ALARGADO">.

Certo: Ex: O incêndio foi na <LOCAL TIPO="ALARGADO">Rua do Padrão</LOCAL>.
Errado: O incêndio foi na <LOCAL TIPO="CORREIO">Rua do Padrão</LOCAL>.
Certo: Eu deixei o carro na <LOCAL TIPO="ALARGADO">Praça da Alegria</LOCAL>.
Certo: Eu moro na <LOCAL TIPO="CORREIO">Praça da Alegria, nº 7</LOCAL>.
Errado: Eu moro na <LOCAL TIPO="ALARGADO">Praça da Alegria, nº 7</LOCAL>.

Diferença entre GEOGRAFICO e ADMINISTRATIVO

Fazemos uma diferença clara entre acidentes geográficos (naturais, objecto de estudo da geografia física) e localizações de geografia humana. Amazónia é um local GEOGRAFICO, Brasil é um local ADMINISTRATIVO. Nos casos em que existe uma coincidência exacta, como é o caso de por exemplo a Islândia, usa-se o tipo ADMINISTRATIVO.

9 Categoria OBRA

A categoria OBRA refere-se a qualquer coisa feita pelo Homem e que tenha um nome próprio (não comum).

9.1 Tipo REPRODUZIDA

Obras das quais há muitos exemplares, o nome representa o original a partir do qual se fazem as reproduções ("Turn it on again", "Olhai os Lírios do Campo", "E Tudo o Vento Levou", "Sinfonia em si bemol", de Carlos Seixas, Bíblia).

Certo: O álbum de música rock mais famoso é o
'<OBRA TIPO="REPRODUZIDA">Achtung Baby</OBRA>'.

9.2 Tipo ARTE

Obras ou objectos das quais há um exemplar único, tais como *Torre Eiffel*, *Guernica*, *Cristo-Rei*, *Capela Sistina*, *Igreja da Luz*, *Ponte da Arrábida*.

Certo: O <OBRA TIPO="ARTE">Mosteiro dos Jerónimos</OBRA> é o expoente máximo do estilo manuelino.

'Arte' também como LOCAL

De reparar que, no caso anterior, onde a EM se refere a certos edifícios ou monumentos, o seu contexto pode ser a sua localização ou a obra em si. (Por exemplo, *Igreja da Luz* – LOCAL ou OBRA?).

Certo: A <OBRA TIPO="ARTE">Igreja da Luz</OBRA> tem um estilo único.
Certo: Encontramo-nos amanhã ao pé da <LOCAL TIPO="ALARGADO">Igreja da Luz</LOCAL>.

9.3 Tipo PUBLICACAO

Este tipo abrange obras escritas não referidas pelo nome, tais como citações de livros, artigos, decretos, directivas, entre outros. A etiqueta deve abranger todas as palavras relacionadas com a publicação, inclusivé nomes de editoras e/ou locais da publicação (Exemplos: *Maia et al. (2004)*, *Santos & Sarmento (2003:114)*, *Mota (op.cit.)*, *Decreto Lei 254/94*).

Certo: O <OBRA TIPO="PUBLICACAO">Decreto Lei nº 31/3 de 2005</OBRA> diz que isso é proibido.

Certo: Os resultados foram semelhantes aos produzidos por <OBRA TIPO="PUBLICACAO">(Santos et al, 2005)</OBRA>.

Citações a publicações no texto

O tipo PUBLICACAO engloba apenas produtos literários que são referidos por citações no texto. Quando se refere uma obra conhecida, é usada o tipo REPRODUZIDA.

Certo: <OBRA TIPO="REPRODUZIDA">Os Lusíadas</OBRA> descrevem a odisseia dos portugueses.

Certo: <OBRA TIPO="PUBLICACAO">Camões(1554)</OBRA> diz que...

Referências à obra ou estilo de um autor

Quando se refere a obra de um autor pelo nome do autor, mencionando um estilo ou a totalidade do seu trabalho, deve-se marcar como <ABSTRACCAO TIPO="OBRA1">, e não como <PESSOA TIPO="INDIVIDUAL"> ou <OBRA TIPO="PUBLICACAO">.

Certo: Em <ABSTRACCAO TIPO="OBRA">Camões</ABSTRACCAO>, as musas são gregas.

10 Categoria ABSTRACCAO

Esta categoria exprime uma quantidade de ideias que são mencionadas por um nome próprio em português, que nos parecem também relevantes para um sistema de REM.

A categoria engloba áreas do conhecimento e práticas (DISCIPLINA), estados e funções (ESTADO), correntes de pensamento e facções (ESCOLA), planos e projectos (PLANO), marcas (MARCA), ideias abstractas (IDEIA) e os próprios nomes (NOME).

10.1 Tipo DISCIPLINA

Engloba disciplinas científicas, teorias, tecnologias e práticas, tais como *Inteligência Artificial, Neurofisiologia, Teoria da Relatividade, GSM, Tai-Chi, Futebol de 5, Java*.

Também inclui especialidades e áreas de governação, quando citadas como tal (pasta dos Negócios Estrangeiros, ministro/secretário de Estado do Interior).

Certo: O Dr. Silva foi demitido da pasta da <ABSTRACCAO TIPO="DISCIPLINA">Economia</ABSTRACCAO>.

Certo: Este programa foi escrito em <ABSTRACCAO TIPO="DISCIPLINA">Java</ABSTRACCAO>.

10.2 Tipo ESTADO

Engloba estados físicos, condições ou funções, tais como *doença de Alzheimer, AIDS, síndrome de Chang, Sistema Nervoso Central*. As EM de tipo ESTADO devem incluir os prefixos que os tornam estados no seu contexto (por exemplo, *mal de, estado de, doença de, síndrome de*), mesmo que apresentem minúscula.

Certo: As vacas podem apanhar a <ABSTRACCAO TIPO="ESTADO">doença de Creutzfeldt-Jakob</ABSTRACCAO>.

Errado: As vacas podem apanhar a doença de <ABSTRACCAO TIPO="ESTADO">Creutzfeldt-Jakob</ABSTRACCAO>.

10.3 Tipo ESCOLA

Compreende escolas, modas, facções, seitas, entre outros. Exemplos são *Barroco, Renascimento, Bushismo, Testemunhas de Jeová, Darwinismo*.

Certo : O <ABSTRACCAO TIPO="ESCOLA">Nazismo</ABSTRACCAO> surgiu na Alemanha.

10.4 Tipo MARCA

Compreende referências a marcas de produtos e raças de animais. Esta categoria pretende identificar as marcas como um conceito abstracto, como é o caso de a menção a uma marca sugerir credibilidade ou desconfiança.

Certo: O meu cão é um <ABSTRACCAO TIPO="MARCA">Rotweiller</ABSTRACCAO>.

Certo: A <ABSTRACCAO TIPO="MARCA">Vista Alegre</ABSTRACCAO> é reputadíssima no estrangeiro.

Certo: Os <ABSTRACCAO TIPO="MARCA">Toyota</ABSTRACCAO> inspiram confiança.

Errado: O João vende <ABSTRACCAO TIPO="MARCA">Toyotas</ABSTRACCAO>.

Certo: O João vende <COISA TIPO="MEMBROCLASSE">Toyotas</COISA>.

Note-se que, no último caso, a EM *Toyota*, que foi classificada como <COISA TIPO="MEMBROCLASSE">, refere os produtos, e não a marca.

10.5 Tipo PLANO

Abrange medidas políticas, administrativas e/ou financeiras, assim como projectos ou acordos, que são designadas por um nome único (*Plano Marshall, Orçamento Larou, Rendimento Mínimo Garantido*).

Certo: O <ABSTRACCAO TIPO="PLANO">Pacto de Varsóvia</ABSTRACCAO> proibiu o comércio da Polónia com o Ocidente.

O exemplo anterior considera *Pacto de Varsóvia* no seu contexto de acordo ou medida política. No entanto, o mesmo nome pode definir uma ORGANIZACAO ou, até, uma efeméride:

Certo: Os países do <ORGANIZACAO TIPO=ADMINISTRACAO>Pacto de Varsóvia</ORGANIZACAO> desenvolveram uma política comum.

Certo: O <ACONTECIMENTO TIPO="EFEMERIDE">Pacto de Versóvia</ACONTECIMENTO> comemora 40 anos de idade.

10.6 Tipo IDEIA

As ideias ou ideais são muitas vezes EM que representam conceitos abstractos, mas que são normalmente referenciados por outros conceitos mais concretos, como é o exemplo de:

Certo: A honra da <ABSTRACCAO TIPO="IDEIA">França</ABSTRACCAO> estava em jogo.

Neste exemplo, o conceito abstracto é a honra, retirado a partir da referência *França*.

A candidatura para a <ABSTRACCAO|ORGANIZACAO TIPO="IDEIA|ADMINISTRACAO">Presidência da República</ABSTRACCAO|ORGANIZACAO>

Neste caso, pode-se interpretar *Presidência da República* (note-se que não é um CARGO, CARGO seria *Presidente da República*) como uma organização, mas também pode-se interpretar a EM como uma referência a um órgão de poder, um conceito mais abstracto do que a ORGANIZACAO.

10.7 Tipo NOME

Por vezes uma dada EM está a representar apenas o nome, e como tal deve ser identificada como um NOME.

Certo: Achei um cão. Vou dar-lhe o nome de <ABSTRACCAO TIPO="NOME">Bobi</ABSTRACCAO>.

Certo: O magnata criou uma empresa chamada <ABSTRACCAO TIPO="NOME">Cauca7</ABSTRACCAO>.

10.8 Tipo OBRA

Quando a referência a um autor pressupõe um estilo artístico ou o seu trabalho artístico, deve ser de categoria <ABSTRACCAO TIPO="OBRA">:

Certo: Em <ABSTRACCAO TIPO="OBRA">Camões</ABSTRACCAO>, as musas são gregas.

11 Categoria VALOR

Valores, como o nome indica, podem referir-se a quantidades absolutas ou relativas (QUANTIDADE), designar dinheiro (MOEDA) ou classificações desportivas, ordinais normais e outras (CLASSIFICACAO). Os itens numéricos a marcar ordem no texto não são considerados EM.

Quando há uma referência a um intervalo de valores, os seus limites devem ser etiquetados como duas EM distintas, e não como uma única EM, como ilustra o caso abaixo:

Certo: Entre <VALOR TIPO="QUANTIDADE">7</VALOR> a <VALOR TIPO="QUANTIDADE">10 metros</VALOR>.

Errado: Entre <VALOR TIPO="QUANTIDADE">7 a 10 metros</VALOR>.

11.1 Tipo CLASSIFICACAO

Engloba valores que traduzem classificação, ordenação ou pontuação (Exemplos: 2-0, 15', 3^a). Enumerações de parágrafos, tópicos e outras secções não devem ser etiquetados.

Tempos como medida de classificação

No exemplo anterior, a EM 15' só é uma classificação quando designa um tempo pelo qual se mede uma dada competição:

Certo: 1º lugar - Ferrari, com o tempo de <VALOR TIPO="CLASSIFICACAO">3' 57' '</VALOR>.

Errado: O golo foi apontado aos <VALOR TIPO="CLASSIFICACAO">14' </VALOR> por Deco.

Certo: O golo foi apontado aos <VALOR TIPO="QUANTIDADE">14' </VALOR> por Deco.

Números como pontuação

Muitas vezes, os números também pertencem a classificações, e como tal, devem ser etiquetados como tal.

Certo: Classificação: <VALOR TIPO="CLASSIFICACAO">1º</VALOR> FC Porto, <VALOR TIPO="CLASSIFICACAO">89</VALOR> pontos.

Números ordinais de eventos organizados

Expressões numéricas incluídas no nome de um evento ou de um cargo não são de categoria VALOR, como ilustram os seguintes exemplos:

Certo: Vai abrir a <ACONTECIMENTO>6ª Exposição Mundial de Cinema <ACONTECIMENTO>.

Errado: Vai abrir a <ACONTECIMENTO><VALOR TIPO="CLASSIFICACAO">6ª</VALOR> Exposição Mundial de Cinema<ACONTECIMENTO>.

Certo: <PESSOA TIPO="CARGO" MORF="M,S">33º Governador da Califórnia</PESSOA>.

Errado: <VALOR TIPO="CLASSIFICACAO">33º</VALOR> <PESSOA TIPO="CARGO" MORF="M,S">Governador da Califórnia</PESSOA>.

Graus escolares e académicos

Classificações referentes a anos escolares não devem ser etiquetados. Esta norma estende-se a graus académicos (*Mestrado, Licenciatura*, etc).

Certo: Reprovei na 4ª classe.

Errado: Reprovei na <VALOR TIPO="CLASSIFICACAO">4ª</VALOR> classe.

Certo: Tenho Mestrado em <ABSTRACCAO TIPO="DISCIPLINA">Pecuária </ABSTRACCAO>.

11.2 Tipo MOEDA

Abrange valores monetários (Exemplos: 300\$00, \$US 15, £39, Cr 500, 50 contos, 30 milhões de cruzeiros). A etiqueta deve abranger a unidade monetária, mesmo que esta esteja em minúsculas.

Certo: O carro custou-me <VALOR TIPO="MOEDA">20000 euros</VALOR>.

Errado: O carro custou-me <VALOR TIPO="MOEDA">20000</VALOR> euros.

11.3 Tipo QUANTIDADE

Engloba percentagens, números soltos, e, caso uma quantidade tenha unidades, a própria unidade. Não engloba unidades monetárias, já abrangidas pelo tipo MOEDA. Exemplos: 15 m, 30 kg, 50 mm, 1,4 kHz, 27º C, 23%, 2.500, pH 2,5.

Por unidades entendem-se as usadas para medir propriedades como distância, tempo, luz, área, volume, peso, massa, etc, e não objectos que sejam contados. Como tal, em 6 *pessoas* ou 9 *folhas de papel*, pessoas e folhas de papel não são para ser incluídas na etiqueta.

As unidades podem ser complexas, como em 23 *metros quadrados*, 9m x 6m, 3 *Bar*, 4 *quilos por metro cúbico*. Toda a especificação da quantidade é para ser incluída na EM.

Embora exista uma categoria TEMPO, esta não abrange evidentemente referências a quantidades de tempo, como ilustra o seguinte exemplo:

Certo: Eu tenho <VALOR TIPO="QUANTIDADE">19 anos</VALOR>.
Errado: Eu tenho <VALOR TIPO="QUANTIDADE">19</VALOR> anos.

12 Categoria VARIADO

Esta etiqueta deve abranger outras referências que são relevantes e que cumpram as regras definidas acima para serem consideradas EM, mas que não são abrangidas nas outras categorias.

Exemplos (não exaustivos) são prêmios, fenômenos naturais ou papéis de teatro (*prémio Valmor*, *tufão El Niño*, *voo 714*, *Rei Lear*).

Para simplificar a forma de processar a classificação semântica no HAREM, a categoria VARIADO deve ser obrigatoriamente expressa com o tipo OUTRO.

Certo: Eu recebi o <VARIADO TIPO="OUTRO">Prémio Camões</VARIADO> o ano passado.

A Palavras em minúscula consideradas como parte integrante de EM

categoria PESSOA

senhor, senhora, doutor, doutora, padre, cónego, deputado, chanceler, lorde, subprocurador-geral, presidente, rei, rainha, miss, major, comandante, capitão, brigadeiro, seu, tio, irmã, irmão, mana, mano, prima, primo, avô, avó, pai, mãe

categoria TEMPO

janeiro, fevereiro, março, abril, maio, junho, julho, agosto, setembro, outubro, novembro, dezembro, século, anos

categoria LOCAL (tipo ALARGADO)

número, nº, sala, abreviaturas de nomes de meses ligados por barra (para indicar o volume de uma revista, por exemplo *jan./dez.*)

categoria ABSTRACCAO (tipo ESTADO)

doença, mal, síndrome, estado

categoria ABSTRACCAO (tipo NOME)

Todos os casos descritos em categoria PESSOA.

Referências

- [1] Nuno Cardoso. Avaliação de Sistemas de Reconhecimento de Entidades Mencionadas. Proposta de tese de mestrado. Faculdade de Engenharia da Universidade do Porto. Janeiro de 2006. <http://www.linguateca.pt/documentos/NCardosoPropostaTese.pdf>.
- [2] Nuno Cardoso. Avaliação de Sistemas de Reconhecimento de Entidades Mencionadas. Apresentação no 2º Simpósio Doutoral da Linguateca, FCUL, Lisboa, Portugal, 10–11 de Abril de 2006. <http://www.linguateca.pt/documentos/CardosoSDL2006.pdf>.
- [3] Nuno Cardoso. Avaliação de Sistemas de Reconhecimento de Entidades Mencionadas. Tese de Mestrado, Faculdade de Engenharia da Universidade do Porto, Outubro 2006. Também disponível como Relatório Técnico DI-FCUL TR–06–26.
- [4] Nuno Cardoso, Diana Santos e Rui Vilela. Directivas para identificação e classificação morfológica na colecção dourada do HAREM. 29 de Março de 2006. Republicado como Relatório técnico DI-FCUL TR–06–19.
- [5] Diana Santos. HAREM: the first evaluation contest for Named Entity Recognition in Portuguese. IST, Lisboa, Portugal. 24 de Fevereiro de 2006. <http://www.linguateca.pt/documentos/SantosISTFev2006.pdf>.
- [6] Diana Santos. Reconhecimento de entidades mencionadas. Palestra convidada na PUC, Rio de Janeiro, Brasil, 18 de Maio de 2006. <http://www.linguateca.pt/documentos/SantosPalestraPUCRio2006.pdf>.
- [7] Diana Santos, editora. *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press. 2007.
- [8] Diana Santos e Nuno Cardoso. “A Golden Resource for Named Entity Recognition in Portuguese”. Em Renata Vieira, Paulo Quaresma, Maria das Graças Volpe Nunes, Nuno J. Mamede, Cláudia Oliveira e Maria Carmelita Dias, editores, *Proceedings of the 7th International Workshop on Computational Processing of the Portuguese Language, PROPOR 2006*, volume 3960 de *Lecture Notes in Computer Science*, págs. 69–79, Itatiaia, Brasil, 13–17 Maio 2006. Springer.
- [9] Diana Santos e Nuno Cardoso, editores. *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro*. Linguateca. 2007. Em preparação.
- [10] Diana Santos, Nuno Cardoso e Nuno Seco. “Avaliação no HAREM: Métodos e Medidas”. Relatório Técnico DI-FCUL TR–06–17, Departamento de Informática da Faculdade de Ciências da Universidade de Lisboa, Outubro 2006.
- [11] Diana Santos, Nuno Cardoso, Nuno Seco e Rui Vilela. “Breve introdução ao HAREM”. Em Diana Santos e Nuno Cardoso, editores, *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro*, Linguateca, 2007.
- [12] Diana Santos, Nuno Seco, Nuno Cardoso e Rui Vilela. “HAREM: An Advanced NER Evaluation Contest for Portuguese”. Em Nicoletta Calzolari, Khalid Choukri, Aldo Gangemi, Bente Maegaard, Joseph Mariani, Jan Odjik e Daniel Tapias, editores, *Proceedings of the 5th International Conference on Language Resources and Evaluation, LREC 2006*, págs. 1986–1991, Génova, Itália, 22–28 Maio 2006. ELRA.
- [13] Nuno Seco, Diana Santos, Nuno Cardoso e Rui Vilela. “A Complex Evaluation Architecture for HAREM”. Em Renata Vieira, Paulo Quaresma, Maria das Graças Volpe Nunes, Nuno J. Mamede, Cláudia Oliveira e Maria Carmelita Dias, editores, *Proceedings of the 7th International Workshop on Computational Processing of the Portuguese Language, PROPOR 2006*, volume 3960 de *Lecture Notes in Computer Science*, págs. 260–263, Itatiaia, Brasil, 13–17 Maio 2006. Springer.

Índice

Nota Preliminar	1
1 Introdução	2
2 Regras gerais de etiquetagem	2
2.1 Recursividade das etiquetas	3
2.2 Vagueza na classificação semântica	3
2.3 Vagueza na identificação	4
2.4 Critérios de identificação de uma EM	4
2.5 Relação entre a classificação e a identificação	5
2.6 Escolha da EM máxima	5
3 Categoria PESSOA	6
3.1 Tipo INDIVIDUAL	6
3.2 Tipo GRUPOIND	7
3.3 Tipo CARGO	7
3.4 Tipo GRUPOCARGO	7
3.5 Tipo GRUPOMEMBRO	7
3.6 Tipo MEMBRO	8
4 Categoria ORGANIZACAO	8
4.1 Tipo ADMINISTRACAO	8
4.2 Tipo EMPRESA	9
4.3 Tipo INSTITUICAO	9
4.4 Tipo SUB	10
5 Categoria TEMPO	11
5.1 Tipo DATA	11
5.2 Tipo HORA	12
5.3 Tipo PERIODO	12
5.4 Tipo CICLICO	13
6 Categoria ACONTECIMENTO	13
6.1 Tipo EFEMERIDE	13
6.2 Tipo ORGANIZADO	13
6.3 Tipo EVENTO	14
7 Categoria COISA	14
7.1 Tipo OBJECTO	14
7.2 Tipo SUBSTANCIA	15
7.3 Tipo CLASSE	15
7.4 Tipo MEMBROCLASSE	15
8 Categoria LOCAL	16
8.1 Tipo CORREIO	16
8.2 Tipo ADMINISTRATIVO	16
8.3 Tipo GEOGRAFICO	17
8.4 Tipo VIRTUAL	17

8.5	Tipo ALARGADO	18
9	Categoria OBRA	18
9.1	Tipo REPRODUZIDA	19
9.2	Tipo ARTE	19
9.3	Tipo PUBLICACAO	19
10	Categoria ABSTRACCAO	20
10.1	Tipo DISCIPLINA	20
10.2	Tipo ESTADO	20
10.3	Tipo ESCOLA	20
10.4	Tipo MARCA	21
10.5	Tipo PLANO	21
10.6	Tipo IDEIA	21
10.7	Tipo NOME	22
10.8	Tipo OBRA	22
11	Categoria VALOR	22
11.1	Tipo CLASSIFICACAO	22
11.2	Tipo MOEDA	23
11.3	Tipo QUANTIDADE	23
12	Categoria VARIADO	24
A	Palavras em minúscula consideradas como parte integrante de EM	25
	Referências	26